RELATIONSHIPS BETWEEN FITNESS REPORT RATINGS AND EXPERIMENTAL RATINGS OF JOB PERFORMANCE AND POTENTIAL

25X1 [redacted]

Prepared in Conjunction with the Task Force on Systems Analysis of Psychological Data Pertaining to Career Trainees

Office of Medical Services
Assessment and Evaluation Staff
July 1969

SUMMARY

Fitness Report Ratings of performance were compared with specially-devised ratings of job performance and potential gathered by the Committee on Professional Manpower (referred to as Manpower Ratings). A group of 743 young Agency Professionals (384 CTs and 359 non-CTs) were studied, for whom Fitness Report Ratings and Manpower Ratings were available. Moderate-sized relationships were found between the two systems of performance evaluation. The size of these relationships was approximately the same for both CTs and non-CTs. Fitness Report Ratings were found to be as highly related to Manpower Ratings of potential as they were to Manpower Ratings of performance (after appropriate statistical corrections were made for difference in the scales). This suggests that Fitness Report Ratings reflect supervisors' estimates of performance and potential to about the same degree. Despite the fact that the Manpower Ratings were not shown to the persons who were rated, while the Fitness Report Ratings were shown, it was found that the mean (average) level of the Manpower Ratings of Overall Performance was essentially the same as the mean level of the Fitness Report Ratings of Overall Performance. The Manpower Ratings of Overall Performance, however, resulted in much greater variability (spread) of ratings than was found in Fitness Report Ratings. The Fitness Report System, as it presently is used, is essentially a 2-point rating scale with approximately 95% of all persons receiving a rating of either "Strong" or "Proficient." The Manpower Ratings of Overall Performance provided a middle category between "Strong" and

"Proficient" and another between "Strong" and "Outstanding" with the result that each of four categories contained 15% or more of the total group of people who were rated.

RELATIONSHIPS BETWEEN FITNESS REPORT RATINGS AND EXPERIMENTAL RATINGS OF JOB PERFORMANCE AND POTENTIAL

The purpose of this study was to compare the job performance ratings produced by the Agency's Fitness Reporting System (Form 45) with the specially-devised ratings of job performance and potential gathered by the Committee on Professional Manpower.[1] By comparing the Fitness Report Ratings individuals receive with the ratings they receive on experimental ratings made under more "ideal" circumstances (such as those gathered by the Committee on Professional Manpower), it becomes possible to learn certain things about the Agency's Fitness Reporting System -- e.g., how effectively Fitness Reports differentiate among people, whether the adjectives used in Fitness Reports to describe employees' performances ("Strong," "Proficient," etc.) are the same adjectives that would be used to describe these people's performances if the Fitness Reports did not have to be shown to the persons being rated and the degree of relationship between the existing Fitness Reporting System and experimental ratings of job performance and future potential.

METHOD

Three hundred eighty-four male CTs and 359 male non-CTs who were included in the survey of the Committee on Professional Manpower were included in this study. These young professionals had entered on duty during Fiscal Years 1963 through 1967 at grades GS-07 through GS-12. In response to a request by the Committee, the total group of 743 employees were rated by their immediate

---

[1]The Committee on Professional Manpower was established by the Executive Director in late 1967 to "examine the quality of recently-appointed junior professional officer personnel in the Agency."

supervisors on six dimensions of current job performance and future job potential in January, 1968 (these six ratings are hereafter referred to as Manpower Ratings). Three ratings of job performance were made for each individual in the sample: his overall performance, the quantitative aspects of his performance (i.e., how much work he gets done), and the qualitative aspects of his performance (i.e., the quality of his work). These three ratings took the form of 7-point scales which were very similar to the traditional WAPSO system, but with two additional categories ("Outstanding," "Between Outstanding and Strong," "Strong," "Between Strong and Proficient," "Proficient," "Adequate," and "Weak"). Three separate ratings of future potential were also produced for each person. First, the supervisor rated overall potential on a 5-point scale ("Outstanding," "Above Average," "Average," "Below Average," "Weak"). Next, each supervisor predicted (on a "yes-no" scale) whether his supervisee had the potential to eventually reach a senior level (GS-15) position in the Agency. Finally, each supervisor predicted whether his supervisee would eventually attain supergrade status (GS-16) in the Agency.

These Manpower Ratings differed from the conventional Fitness Report Ratings in several respects. As already mentioned, they were not limited to current job performance -- three ratings were designed to tap the supervisor's estimate of each individual's future job potential, including his advancement potential. On the measures of current job performance, two additional categories ("Between Outstanding and Strong" and "Between Strong and Proficient") were

added to the WAPSO Rating Scale in the hope that these two extra categories might encourage finer discriminations among the performances of people. Perhaps the major difference between these Manpower Ratings and conventional Fitness Report Ratings had to do with the way they were processed. Unlike Fitness Report Ratings, which are normally seen by the persons being rated, the Manpower Ratings were not shown to the individuals being rated, nor were they to be included in any official records. Thus, it may be presumed that these Manpower Ratings were "purer" measures than the Fitness Report Ratings since they were made for research purposes only with no apparent need on the part of the supervisors to "slant" them for any reason.

Fitness Report Ratings were obtained on the 743 employees on whom Manpower Ratings were available.[1] It was decided to obtain only the "Overall Performance" rating from each Fitness Report, since the number of specific duties which are rated varies from individual to individual, and in addition, the specific duty ratings are not available in computer storage. For each individual in the study, the Overall Performance Ratings from his three most recent Fitness Reports were obtained. Then, the Fitness Report Rating falling closest in time to January, 1968 (the month in which the Manpower Ratings were made) was selected for each person to serve as the single rating to which his Manpower Ratings would be compared. Since for some individuals,

---

[1]Appreciation is expressed to [redacted] for their assistance in obtaining Fitness Report Ratings.

25X1

there was a considerable interval between the time Fitness Report Ratings and Manpower Ratings were made, it was decided to divide the groups of CTs and non-CTs into three "Proximity Groups" based on the time span between, and the sequence of, the two sets of ratings. Table 1 presents the number of persons falling into each of these Proximity Groups. Two-thirds of the total group of 743 employees fell into Proximity Group I (Fitness Report Ratings made anywhere from the same month as to four months before the Manpower Ratings were obtained). Since it was more likely that the same supervisors would have prepared both the Fitness Report Ratings and the Manpower Ratings for those employees in Proximity Group I than for those employees in Groups II (Fitness Report Ratings made five to 16 months before Manpower Ratings) and III (Fitness Report Ratings made one to eight months after Manpower Ratings), and since a major objective of this study was to compare these two types of ratings under as comparable conditions as possible, the bulk of the analyses that follow will focus upon the individuals in Proximity Group I.

RESULTS

Relationships Between Fitness Report Ratings and Manpower Ratings

Table 2 presents the correlations between the six Manpower Ratings and the Overall Performance Ratings from Fitness Reports for CTs and non-CTs in the three Proximity Groups. Close inspection of this table suggests a number of trends. First of all, the correlations between Fitness Report Ratings and

TABLE 1

A DESCRIPTION OF THE CT AND NON-CT SAMPLES BASED UPON THE TIME SPAN BETWEEN FITNESS REPORT RATINGS AND MANPOWER RATINGS

	Proximity Group I (FR Ratings Closely Preceded Manpower Ratings -- FR Ratings Gathered Between 9/67 and 1/68)	Proximity Group II (FR Ratings Farther in Time before Manpower Ratings -- FR Ratings Gathered Between 9/66 and 8/67)	Proximity Group III (FR Ratings Gathered after Manpower Ratings: FR Ratings Gathered Between 2/68 and 9/68)	TOTALS
CTs	260	58	66	384
Non-CTs	235	44	80	359
TOTALS	495	102	146	743

TABLE 2

CORRELATIONS[1] BETWEEN THE SIX MANPOWER RATINGS AND THE OVERALL PERFORMANCE RATING FROM FITNESS REPORTS FOR CTS AND NON-CTS IN THE THREE PROXIMITY GROUPS[2]

CORRELATIONS WITH FITNESS REPORT "OVERALL PERFORMANCE"

	CTS				NON-CTS			
Manpower Rating	All CTs	Group I	Group II	Group III	All Non-CTs	Group I	Group II	Group III
Overall Performance	.52	.55	.53	.37	.53	.57	.47	.50
Quantitative Performance	.45	.47	.39	.39	.50	.55	.42	.45
Qualitative Performance	.49	.55	.49	.27	.45	.48	.50	.41
Senior Level Potential	.30	.34	.17	.28	.34	.31	.44	.37
Supergrade Potential	.30	.31	.34	.26	.27	.30	.20	.27
Overall Potential	.42	.44	.47	.32	.43	.40	.45	.51

[1]Correlation coefficients can range from -1.00 to +1.00. A coefficient of -1.00 indicates a perfect negative relationship, +1.00 indicates a perfect positive relationship, and .00 indicates no relationship whatsoever.

[2]The three Proximity Groups and the sample sizes are defined in Table 1.

Manpower Ratings are only moderate. Even for the CTs and non-CTs in Proximity Group I, whose Fitness Reports were written shortly before their Manpower Ratings were made, the highest correlations were only in the middle .50's, meaning that a sizeable percentage of the persons rated did not receive the same (or highly similar) ratings on the two types of ratings. Secondly, the pattern and size of the relationships between Fitness Report Ratings and Manpower Ratings were essentially the same for the CT and non-CT Groups. Thirdly, the Manpower Ratings of performance (Overall Performance, Quantitative Performance, and Qualitative Performance) were more highly related to Fitness Report Ratings of Overall Performance than were the Manpower Ratings of potential (Senior Level Potential, Supergrade Potential, and Overall Potential). This finding seems logical enough; Fitness Report Ratings are intended to be measures of current performance and hence should correlate more highly with the Manpower Ratings of performance than with the Manpower Ratings of potential. However, it will be recalled that all three of the Manpower Ratings of performance were 7-point scales, while the Manpower Ratings of potential were either 5-point scales (in the case of Overall Potential) or 2-point scales (in the cases of Senior Level and Supergrade potential). Other things being equal, the fewer the categories in a rating scale, the less will be the variability (spread of ratings) and the lower will be the correlation between that rating scale and an outside criterion (in this case, Fitness Report Ratings). This statistical phenomenon is known as restriction of range;

when corrective formulas are applied which estimate what the correlations would be if the variability of the two sets of ratings were equal, the correlations between the Manpower Ratings of potential and Fitness Report Ratings in Table 2 become as large as the correlations between the Manpower Ratings of performance and Fitness Report Ratings. Thus, it may be concluded that Fitness Report Ratings are as highly related to the Manpower Ratings of potential as they are to the ratings of performance, when the Manpower Ratings of Potential are corrected for restriction of range.

Inspection of Table 2 also suggests a tendency for higher correlations between Manpower Ratings and Fitness Reports for those individuals whose Fitness Reports were written within four months before their Manpower Ratings were made (in comparison with those whose Fitness Reports were written more than four months before or from one to eight months after their Manpower Ratings were made). For the CT sample, four of six of the Manpower Ratings were more highly correlated with Fitness Report Ratings in Proximity Group I than in the other two Proximity Groups. For the non-CT sample, three of six of the Manpower Ratings found their highest correlation with Fitness Report Ratings in Proximity Group I. In each of these cases, only two ratings would be expected (by chance alone) to have their highest correlation in Proximity Group I. This type of finding was expected and tends to boost confidence in the obtained correlations between Fitness Ratings and Manpower Ratings; the further separated in time two sets of ratings are from each

other, the more opportunity there is for interpolated events (e.g., changes in supervisors, different job demands, etc.) to lower the correlations between the two sets of ratings.

Another way of presenting the relationship between Manpower Ratings and Fitness Report Ratings is shown in Table 3. This table combines the CT and non-CT samples (in Proximity Group I) and presents the Manpower Ratings of Overall Performance received by individuals falling in each of the WAPSO categories on the Fitness Reports. For instance, of the 325 persons receiving an overall rating of "Strong" on their Fitness Report, 2% received a rating of "Outstanding" on their Manpower Rating of Overall Performance, 21% received a rating of "Between Outstanding and Strong," 46% received a rating of "Strong," 22% received a rating of "Between Strong and Proficient," 9% received a rating of "Proficient," and 1% received a rating of "Adequate." This table illustrates that there is some variation in the Manpower Ratings of Overall Performance received by employees falling in the same WAPSO categories on their Fitness Reports. Nevertheless, the data in Table 2 also show there is a definite tendency for persons receiving high (or low) ratings on their Fitness Reports to receive high (or low) Manpower Ratings as well.

A third way of summarizing the relationship between the two sets of ratings is to determine the percentage of people who were "misclassified" on their Fitness Reports, as defined by their subsequent Manpower Ratings of Overall Performance. There may be many different reasons for such misclassification -- the individuals rated may have changed jobs, acquired different supervisors,

TABLE 3

MANPOWER RATINGS OF OVERALL PERFORMANCE RECEIVED BY 495 AGENCY PROFESSIONALS (CTS AND NON-CTS) FALLING IN THE VARIOUS WAPSO CATEGORIES ON THEIR FITNESS REPORTS[1]

	RATING RECEIVED ON MANPOWER RATING OF OVERALL PERFORMANCE							
Fitness Report Rating -- Overall Performance	Outstanding	Between Outstanding and Strong	Strong	Between Strong and Proficient	Proficient	Adequate	Weak	TOTALS (%'s of Total Sample)
Outstanding	36% (5)	36% (5)	21% (3)	7% (1)				3% (14)
Strong	2% (7)	21% (67)	46% (151)	22% (70)	9% (28)	1% (2)		66% (325)
Proficient	1% (1)	1% (2)	16% (24)	39% (59)	34% (51)	9% (14)	1% (1)	31% (152)
Adequate					25% (1)	75% (3)		1% (4)
Weak								
TOTALS (Percentages of Total Sample)	2% (13)	15% (74)	36% (178)	26% (130)	16% (80)	4% (19)	0% (1)	100% (495)

Note--This sample of 495 represents the total number of CTs and non-CTs whose Fitness Ratings were made from 0-4 months before their Manpower Ratings (Proximity Group I). It will be recalled that the highest relationship between Fitness Ratings and Manpower Ratings was obtained for this group.

[1]The percentages given represent the percentages of the groups falling in each major WAPSO category which received each of the seven Manpower Ratings. Actual numbers of people are given in parentheses.

actually changed their levels of performance, or their supervisors may have changed their minds about their performances (possibly since the Manpower Ratings did not have to be shown to the persons being rated) -- and this type of analysis sheds no light upon the actual reasons underlying these changes in ratings. Nevertheless, this way of looking at the data does provide one readily understandable measure of the relationship between Fitness Report Ratings and Manpower Ratings.

Referring to Table 3, we find that 14 persons received a rating of "Outstanding" on their Fitness Reports. Of these 14, five received "Outstanding" and five received "Between Outstanding and Strong" on their Manpower Ratings of Overall Performance. These ten persons may be said to have received Manpower Ratings which were <u>not</u> inconsistent with the adjective assigned to them in their previous Fitness Report. However, of the 14 who received a rating of "Outstanding" on their Fitness Report, those three who received "Strong" and that one who received "Proficient" on the Manpower Rating may be said to have been "misclassified" -- i.e., they were assigned labels on their Manpower Ratings which were inconsistent with their previous Fitness Report Ratings.

Continuing this type of analysis for all the Fitness Report Rating categories in Table 3, it is found that of the total of 495 people, 84 were "misclassified" on their subsequent Manpower Ratings while 411 (or 83%) received ratings which were not inconsistent with their previous Manpower Ratings. Thus, better than four of every five persons in the sample received Manpower Ratings which were not inconsistent with their Fitness

Ratings. Viewed in this manner, it may be said that a high degree of relationship exists between the Fitness Report Ratings of Overall Performance and the Manpower Ratings of Overall Performance, despite correlation coefficients between the two sets of ratings which were only of moderate size.

Comparisons of the Variance (Spread) and the Mean Levels of the Two Sets of Ratings

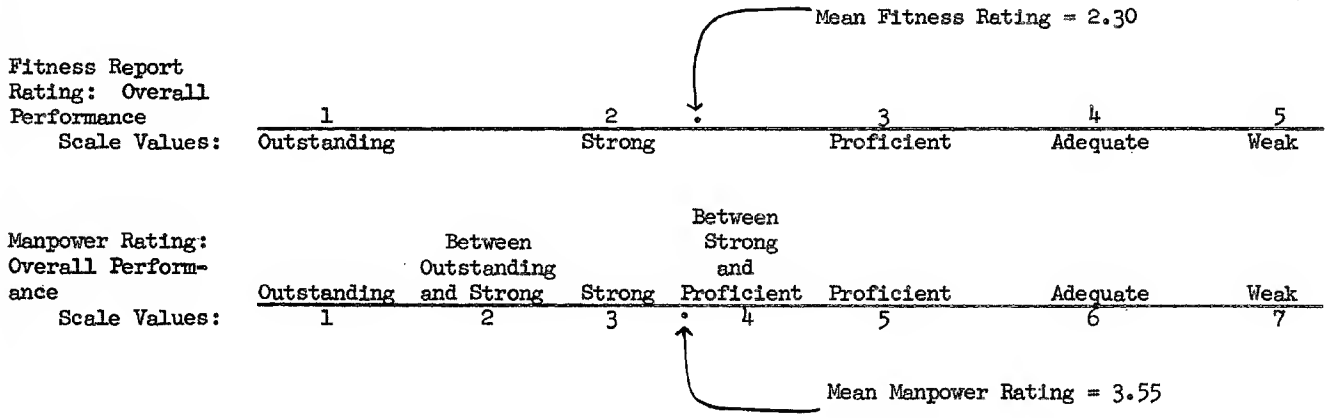
Casual inspection of the marginal totals of Table 3 clearly reveals that there is a great deal more variance (spread) of the Manpower Ratings of Overall Performance than of the Fitness Report Overall Ratings. Sizeable percentages of people (15% or over) fell in four categories on the Manpower Ratings of Overall Performance; for the Fitness Report Ratings of Overall Performance, only two categories contained more than 15% of the total sample. The combined total of these two categories -- "Strong" with 66% and "Proficient" with 31% -- accounted for nearly everyone. In terms of variance defined in statistical terms, the Manpower Ratings of Overall Performance yielded over four times the variance of the Fitness Report Ratings of Overall Performance, a highly significant increase in variance. By placing two additional categories -- "Between Outstanding and Strong" and "Between Strong and Proficient" into the conventional 5-point WAPSO Scale, a great many more distinctions among people were made. It is, of course, possible that part of this increase in variance was due not to the expanded nature of the Manpower Rating Scale, but instead was a result of the Manpower Ratings not being shown to the persons who were

rated, while the Fitness Report Ratings were shown. However, since it is generally known that introduction of additional categories in a scale in the range where cases "pile up" leads to increased variance of that scale, and since it is somewhat unlikely that showing or not showing the ratings to the people being rated would appreciably affect the variance of a scale, it seems reasonable to conclude that most, if not all of, the increase in variance of the Manpower Ratings of performance was due to the addition of two categories.

It could be hypothesized that the conditions under which the Manpower Ratings were made may have caused these ratings to have a lower mean (average) than was obtained on the Fitness Report Ratings for the same individuals. Since the Fitness Report Ratings were shown to the individuals who were rated, while the Manpower Ratings were not, it would seem reasonable to expect that the average of the Fitness Report Ratings would be higher than the average of the Manpower Ratings. Figure 1 provides a comparison of the mean ratings of Overall Performance from Fitness Report Ratings and Manpower Ratings. Since the Fitness Report Ratings were made on a 5-point scale, while the Manpower Ratings were made on a 7-point scale, direct comparison of the actual mean values is meaningless. However, when these mean values are plotted (as in Figure 1) on scales which have been equated on the five adjective points of the WAPSO Fitness Report Scale, it can be seen that the mean levels of the two sets of ratings correspond very closely. For both the Manpower Ratings and the Fitness Report Ratings, the mean rating was slightly below "Strong,"

Figure 1

A COMPARISON OF THE MEAN LEVELS OF THE RATINGS OF OVERALL PERFORMANCE FROM FITNESS REPORTS AND MANPOWER RATINGS (N=384 CTS and 359 NON-CTS)



approximately one-quarter of the way toward "Proficient." Thus, the Manpower Ratings of Overall Performance made for the 743 professionals in the study were not systematically lower than the Fitness Report Ratings for the same group.

Another way of demonstrating the basic similarity between the distributions of Fitness Report Ratings and Manpower Ratings of Overall Performance is presented in Table 4, which shows the percentage of persons whose ratings fell in each of the seven categories of the Manpower Ratings and each of the five categories of the Fitness Report Ratings. In addition, Table 4 shows what the distribution of Manpower Ratings looks like when it is compressed from seven into five categories by apportioning the percentages falling in the two additional categories of the Manpower Ratings ("Between Outstanding and Strong," and "Between Strong and Proficient") into adjacent categories in proportion to the number of persons already in these adjacent categories. Comparing this compressed distribution of Manpower Ratings with the obtained distribution of Fitness Report Ratings reveals a high degree of similarity between the two distributions; this comparison corroborates the finding of no significant difference in the mean levels of the two sets of ratings. Parenthetically, Table 4 clearly demonstrates the increased "spread" of ratings obtained with the 7-point Manpower Ratings; 41% of the total sample fell in the two additional categories of "Between Outstanding and Strong" and "Between Strong and Proficient."

DISCUSSION

The major finding of this study -- a moderately high relationship between

TABLE 4

A COMPARISON OF THE PERCENTAGES OF PEOPLE WHO RECEIVED RATINGS IN EACH OF THE CATEGORIES ON THE MANPOWER RATING OF OVERALL PERFORMANCE AND THE FITNESS REPORT RATING OF OVERALL PERFORMANCE (N=260 CTS AND 235 NON-CTS)

	Outstanding	Between Outstanding and Strong	Strong	Between Strong and Proficient	Proficient	Adequate	Weak
Manpower Rating: Overall Performance	2%	15%	36%	26%	16%	4%	0%
Compressed[1] Manpower Rating: Overall Performance	3%	---	68%	---	24%	4%	0%
Fitness Report Rating: Overall Performance	3%	---	66%	---	31%	1%	0%

NOTE--This sample of 495 represents the total number of CTs and non-CTs whose Fitness Ratings were made from 0-4 months before their Manpower Ratings (Proximity Group I).

[1]These Compressed Manpower Ratings were arrived at by apportioning the percentages falling in the two additional categories of the Manpower Ratings ("Between Outstanding and Strong" and "Between Strong and Proficient") into adjacent categories in proportion to the number of persons already in these adjacent categories.

the experimental Manpower Ratings of Overall Performance and the Fitness Report Rating of Overall Performance -- is encouraging; it suggests that supervisors will say basically the same things about the performance of their supervisees whether or not they are required to communicate these ratings to the persons being rated. This conclusion is further strengthened by the finding of no difference between the average level of Fitness Report Ratings and Manpower Ratings assigned to the same group of 495 individuals. It is, of course, possible that the supervisors in this study were motivated by a desire to be consistent and accordingly assigned Manpower ratings which were in agreement with the Figness Report Ratings they had previously assigned to given individuals. If this motive were present to a large degree, it could lead to results such as those obtained in this study. There is no way of throwing further light on this possibility other than to not require that Fitness Reports be shown to the persons being rated and to record at some later date whether the mean level of Fitness Report Ratings assigned declines.

Perhaps the most striking finding in this study was the increase in the variability (spread) of ratings obtained when only two additional rating categories are added to the WAPSO Rating Scale. As the Fitness Report Scale stands now, it is for all practical purposes a 2-point scale -- only about 5% of those rated receive a rating other than "Strong" or "Proficient." By providing a middle category between "Strong" and "Proficient," and another between "Strong" and "Outstanding," the results of this study suggest that

the Fitness Report Scale could be expanded to contain four effective categories, with greater than 15% of those rated falling in each of these four categories. If it is wished to increase the number of discriminations among people made by the Agency's performance evaluation system, the inclusion of these two additional points on the WAPSO Scale is recommended.